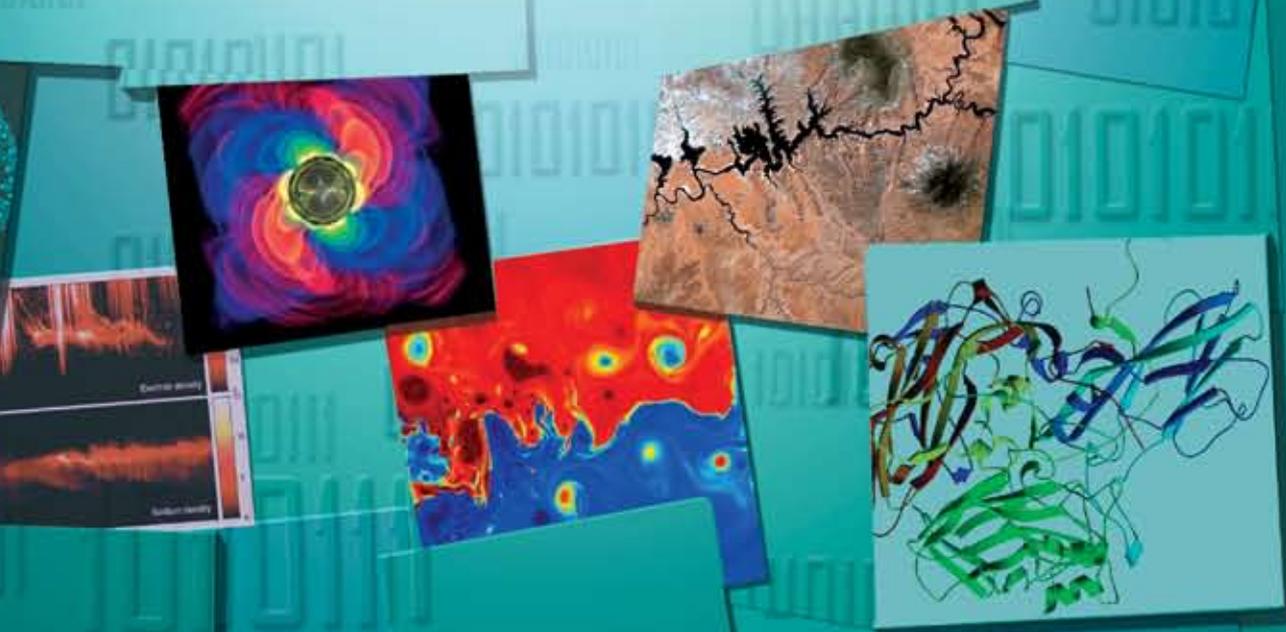


National Science Board

Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century



National Science Foundation

September 2005



Descriptive captions and credits for images on the cover of this report are provided inside the back cover.

National Science Board

**LONG-LIVED DIGITAL DATA
COLLECTIONS: ENABLING
RESEARCH AND EDUCATION
IN THE 21ST CENTURY**

NATIONAL SCIENCE FOUNDATION

SEPTEMBER 2005



NATIONAL SCIENCE BOARD

Terms Expire May 10, 2006

Nina V. Fedoroff, Evan Hugh Professor, Willaman Professor of Life Sciences, and Director, Biotechnology Institute, The Pennsylvania State University, University Park

Jane Lubchenco, Wayne and Gladys Valley Professor of Marine Biology and Distinguished Professor of Zoology, Oregon State University

Diana S. Natalicio, *Vice Chair*, President, The University of Texas at El Paso

Michael G. Rossmann, Hanley Distinguished Professor of Biological Sciences, Purdue University

Daniel Simberloff, Nancy Gore Hunger Professor of Environmental Science, Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville

Warren M. Washington, *Chair*, Senior Scientist and Section Head, National Center for Atmospheric Research

John A. White, Jr., Chancellor, University of Arkansas, Fayetteville

Mark S. Wrighton, Chancellor, Washington University at St. Louis

Terms Expire May 10, 2008

Barry C. Barish, Linde Professor of Physics and Director, LIGO Laboratory, California Institute of Technology

Ray M. Bowen, President Emeritus, Texas A&M University, College Station

Delores M. Etter, ONR Distinguished Chair in S&T, Electrical Engineering Department, U.S. Naval Academy

Kenneth M. Ford, Director, Florida Institute for Human and Machine Cognition

Daniel Hastings, Director, Engineering Systems Division and Professor, Aeronautics and Astronautics and Engineering Systems, Massachusetts Institute of Technology

Elizabeth Hoffman, President, University of Colorado System

Douglas D. Randall, Professor of Biochemistry and Director, Interdisciplinary Program on Plant Biochemistry and Physiology, University of Missouri-Columbia

Jo Anne Vasquez, Mesa, Arizona Public Schools (Retired)

Terms Expire May 10, 2010

Dan E. Arvizu, Director, National Renewable Energy Laboratory

Steven C. Beering, President Emeritus, Purdue University, West Lafayette

G. Wayne Clough, President, Georgia Institute of Technology

Kelvin K. Droegemeier, Regents' Professor & Roger and Sherry Teigen Presidential Professor; Weathernews Chair of Applied Meteorology; Director, Center for Analysis and Prediction of Storms; and Director, Sasaki Institute, University of Oklahoma, Norman

Louis J. Lanzerotti, Distinguished Research Professor, Center for Solar-Terrestrial Research, New Jersey Institute of Technology

Alan Leshner, Chief Executive Officer, American Association for the Advancement of Science

Jon C. Strauss, President, Harvey Mudd College

Kathryn D. Sullivan, President and CEO, Center of Science and Industry

Member *ex officio*

Arden L. Bement, Jr., Director, National Science Foundation

Michael P. Crosby, Executive Officer, National Science Board, and National Science Board Office Director

The National Science Board consists of 24 members plus the Director of the National Science Foundation. Appointed by the President, the Board serves as the policy-making body of the Foundation and provides advice to the President and the Congress on matters of national science and engineering policy.

NATIONAL SCIENCE BOARD
COMMITTEE ON PROGRAMS AND PLANS

Daniel Simberloff, *Chair*

Barry C. Barish

Kelvin K. Droegemeier

Delores M. Etter

Nina V. Fedoroff

Kenneth M. Ford

Louis J. Lanzerotti

Alan Leshner

Jane Lubchenco

Michael G. Rossmann

John A. White, Jr.

Warren M. Washington, *ex officio**

Diana S. Natalicio, *ex officio**

Arden L. Bement, Jr., *ex officio**

* The NSB Chair, NSB Vice Chair, and NSF Director are members *ex officio* of all NSB committees.

[Blank Page]

CONTENTS

Acknowledgments	7
Executive Summary	9
Chapter One: Introduction	13
Chapter Two: The Elements of the Digital Data Collections Universe	17
Overview	17
Individuals and Institutions	17
Data.....	18
Digital Data Collections	20
Example of the Evolution of a Collection: The Protein Data Bank	21
Long-Lived Digital Data Collections	22
Digital Data Common Spaces	23
Conclusions	24
Chapter Three: Roles and Responsibilities of Individuals and Institutions	25
Shared Goals and Responsibilities	25
Data Authors.....	25
Data Managers	26
Data Scientists	27
Data Users	28
Funding Agencies	28
Data Quality Act	29
Chapter Four: Perspectives on Digital Data Collections Policy	31
Overview	31
Need for an Evaluation of NSF Policies	31
Specific Policy Issues	32
1. Proliferating Collections	32
2. Community-Proxy Policy	33
3. Data Sunset and Data Movement	34
4. Data Management Plans	35
5. Data Access/Release Policies	36
6. Digital Data Commons as a Means for Broadening Participation	37
7. Opportunities for Education, Training and Workforce Development	38
8. Duration of NSF Commitment to Support Long-Lived Digital Collections	39
Long-Lived Digital Data Collections and Large Facilities	40

Chapter Five: Findings and Recommendations	43
Workshop Outcomes	43
Recommendations	44
Develop a Clear Technical and Financial Strategy	44
Create Policy for Key Issues Consistent with Technical and Financial Strategy	46
Conclusions	49
Appendices	
A. Task Force Charter	51
B. Sources of Additional Information	53
C. Current Policies on Data Sharing and Archiving	57
D. Digital Data Collections by Categories	73
E. Process for Producing the Report	79
F. Workshop Agenda, November 2003	81
G. Workshop Agenda, March 2004	83
H. Public Comments	85

ACKNOWLEDGMENTS

The National Science Board feels it is especially important to highlight the vital role that Dr. Anita K. Jones, Quarles Professor of Engineering and Applied Science, University of Virginia, played in the genesis of this Board effort, and thank her for her significant contributions. Dr. Jones is a former member of the Board (1998-2004), and served as chair of the Committee on Programs and Plans (CPP) (2002-2004). Dr. Jones provided the initial inspiration and vision for this Board study. She was directly involved with creating the initial draft report and gave the intellectual guidance and leadership that carried the report through final refinement and approval action after her term of appointment to the Board ended.

Dr. Chris Greer, NSF Program Director, Division of Biological Infrastructure, served as Executive Secretary to the CPP Task Force on Long-Lived Digital Data Collections, and deserves special recognition for his direct support and critical intellectual contributions to the overall project from its inception to its completion.

The National Science Board thanks the many members of the science, engineering, education, and policy communities for the generous contribution of their time and intellect to the Board's study. Those who participated as speakers and panelists in the two workshops that informed this report are listed in the agendas for those workshops in Appendices F and G. The names of organizations and individuals who commented on the draft released by the Board for public comment are listed in Appendix H.

We are grateful to the many National Science Foundation staff members who have provided insights and support throughout this study. These include Dr. Janice Earle, Dr. Gerald Guala, Dr. Clifford Jacobs, Dr. Nigel Sharp, Dr. Daniel Newlon, and Dr. Manfred Zorn.

We want to express our appreciation to the National Science Board Office staff for their many essential contributions to this effort, often behind the scenes. Those deserving special note are Ms. Amanda Slocum and Ms. Clara Englert, who provided valuable assistance in the process of inviting and receiving public comments, and Dr. Robert Webber, for his efforts in managing report drafts and coordinating the publication process. We are especially grateful to Dr. Michael P. Crosby, Executive Officer of the National Science Board, for his contributions and insights, as well as for his guidance in navigating the many steps leading to publication of this report.

[Blank Page]